Using XML Notepad to Read, Edit, and Parse FGDC-CSDGM XML Metadata.

Currently many tools exist that allow a user to work with XML metadata files. However, most of the main metadata creation/editing tools do not provide a means of validating a metadata record for compliance with the FGDC-CSDGM standard or its variants, such as the Biological Data Profile (BDP) standard.

One of the most reliable ways to validate metadata files (that is, check files for completion and/or errors) is with the USGS Metadata Parser (MP) utilities. The MP tool and the other tools distributed in the package were developed by Peter Schweitzer and are freely available online (http://geology.usgs.gov/tools/metadata/tools/doc/mp.html). Once configured properly, MP can process metadata files and be used to produce a text file with a list of all errors found. A user can then use this list of errors to find and correct the problems using a text editor or the metadata editor of their choice.

The challenge with this method is that finding and correcting the error within a metadata editor based solely on the error message from MP can be confusing and frustrating. Additionally if there are many errors or complex problems to correct, the process will require multiple iterations of running MP and subsequently correcting errors in a metadata editor. All in all, the process can be fairly time consuming and onerous, especially for those without a detailed understanding of the FGDC-CSDGM standard.

The methodology for metadata editing and validation described in this document tries to address some of these problems with an alternative approach. Using a free XML editor and a detailed XML validation schema provides a means to edit FGDC-CSDGM records much more efficiently.
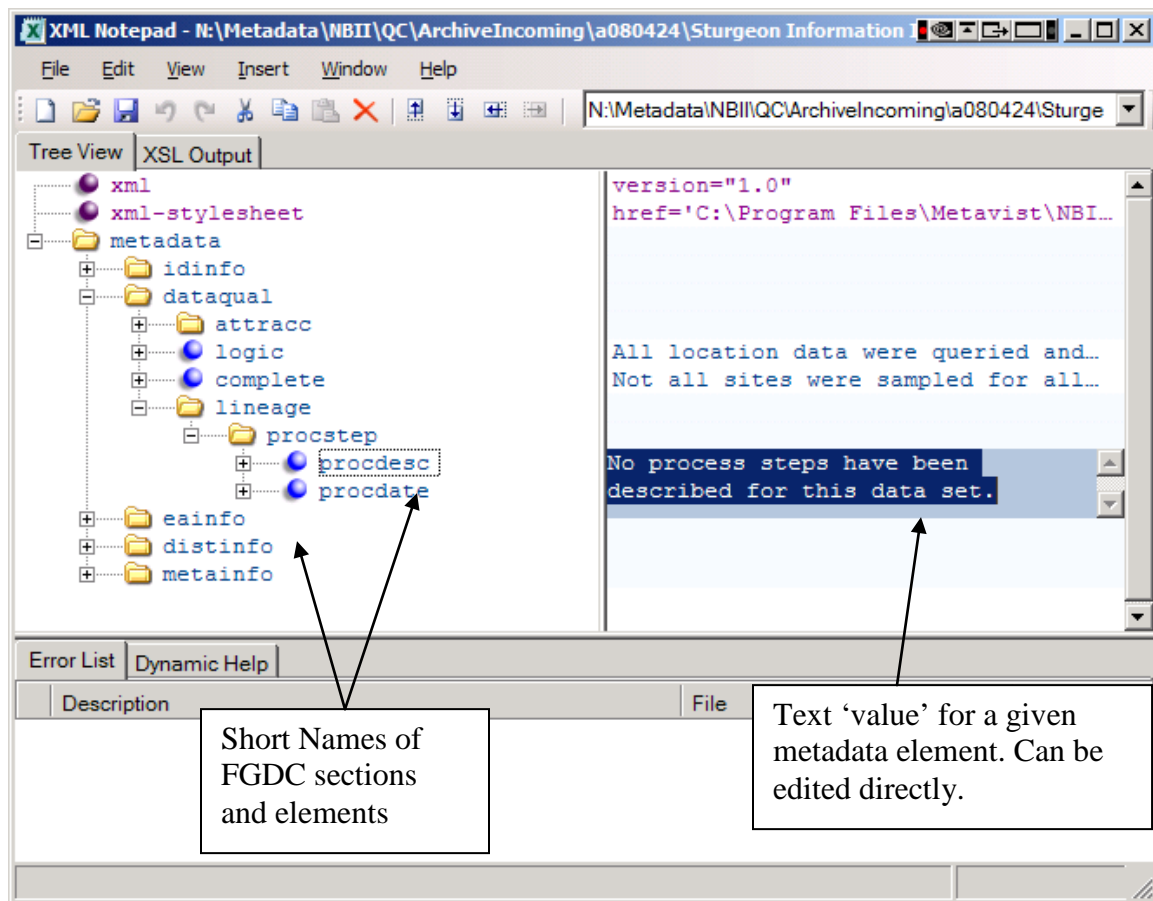
This methodology requires:
1. XML Notepad – Free tool distributed by Microsoft, available at
http://www.microsoft.com/en-us/download/details.aspx?id=7973

2. FGDC .xsd schema files, available here: http://www.fgdc.gov/schemas/metadata/
This schema is also provided with the materials.

3. Biological Data Profile (BDP) FGDC-CSDGM .xsd schema files, if this extension is being used. This schema is also provided with the materials.

4. (Optional) An .xsl stylesheet is nice to have for viewing metadata more easily. One is provided in the materials.

Steps:

1. Install XML Notepad on your machine.

2. Open an XML Metadata record in XML Notepad

On the left you will notice an expandable tree structure with the various elements of your metadata record. The name of each is the FGDC "shortname". The elements that have a folder icon are complex elements that contain additional elements. The elements that have a sphere icon are the elements that hold pieces of information (i.e., text or numerical values). On the right the values of the elements are shown. For longer fields you can click on a value to see the complete entry. By clicking into the field on the right, a user can edit the content associated with a particular element.

Short Names of FGDC sections and elements

Text 'value' for a given metadata element. Can be edited directly.

3. Validate the record against an XSD schema

In order to check that the metadata record is fully compliant with the FGDC/BDP standard it must be run past a schema that checks for missing/misplaced elements or unexpected/non-permitted values within an element. To apply the schema, click on 'View' in the header menu and then 'Schemas…' in the dropdown menu. The box shown below will pop open. Click on the box to the right with the three dots to navigate to the appropriate XSD file.

For FGDC metadata records, this is the '**fgdc-std-001-1998.xsd**' file.

For BDP metadata records, this is the '**BDPfgdc-std-001-1998.xsd**' file.

Click here to navigate to the correct schema
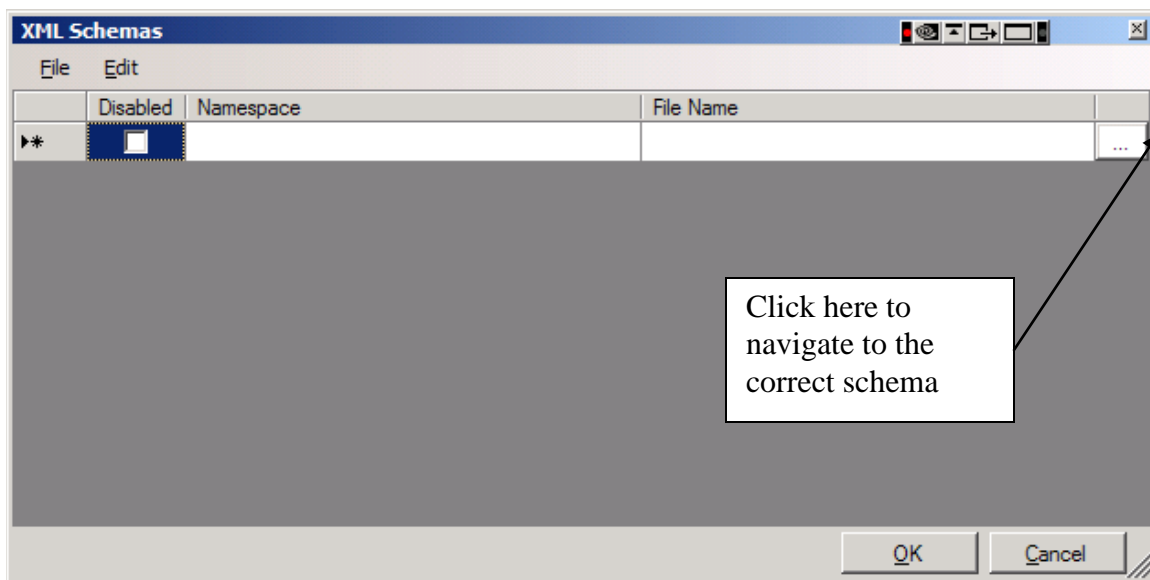
It may be necessary to first remove all previous schemas and then navigate to them again before they will work properly. Also note that, somewhat counter-intuitively, a user must 'Disable' all but the one schema that they wish to apply in XML Notepad. Only one schema can be applied at a time.

Be aware that a schema consists of multiple files. For the schema to work properly, each of the constituent .xsd files must be properly referenced with a valid system path or location so that XML Notepad can locate and apply it.

Although only the one 'top-level' .xsd file needs to selected as the specified schema in XML notepad, this file needs to properly 'point to' the other .xsd files the define the individual sections of the FGDC-CSDGM standard.

Setting up a schema

The screen capture below shows an example of what a complete schema looks like on a user's file system. Notice that there is one top-level .xsd file, and 10 other component files, one for each section of an FGDC-CSDGM record.
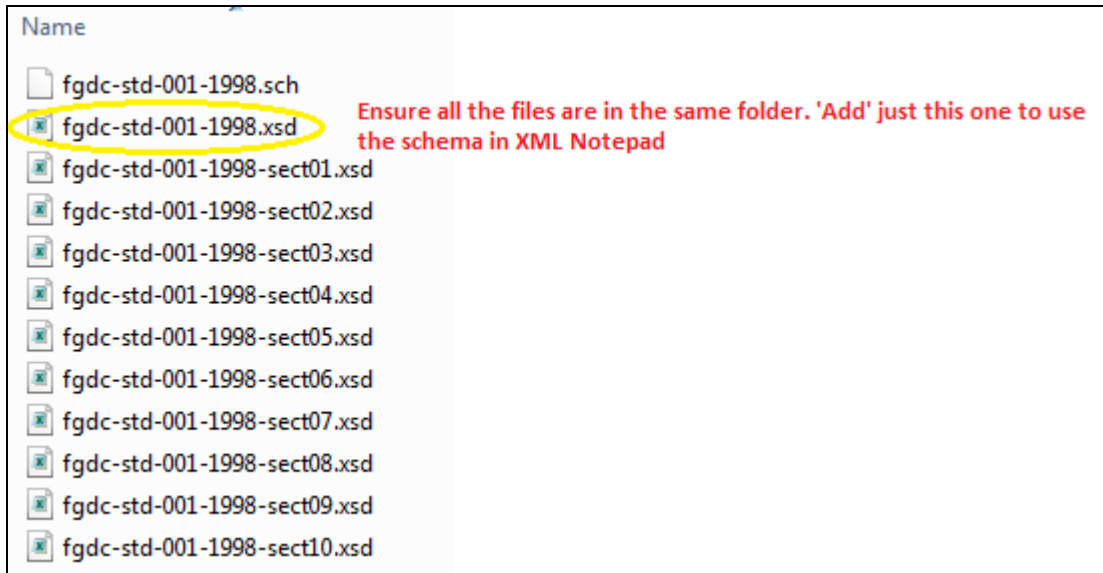
After ensuring that all the necessary component .xsd files are present, a user should point to the 'top-level' .xsd file (shown above, circled in yellow) as the schema to 'Add' from within XML Notepad.

If you have trouble getting the schema to apply properly, please see the **Appendix 1. – Applying/Reconfiguring Schemas in XML Notepad. This issue will be apparent if you see an error that says, "Schema Location Unresolved." In this scenario, dynamic error checking won't be possible until the location of the schema on a user's system has been properly specified. **

4. Find and correct errors

Once the schema is in place a list of all the errors in your metadata file will populate the Error List at the bottom of the application. Clicking on F4 will highlight and move to the next error in the list. If a required element is missing, highlight the parent element and click control Insert. This adds the missing element and shows a list of the names that element could take. Click on the appropriate one (usually there will only be one), then add an appropriate value in the right panel. For help on permissible values, expected element ordering, and general FGDC-CSDGM requirements please see: http://www.fgdc.gov/csdgmgraphical/index.html or download and print the official CSDGM workbook, which can be acquired from here: http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf

Once you are satisfied with the correction click F4 to move onto the next error.

If the element is correct, but has a value that is not valid, clicking in the value box on the right will bring up a dropdown menu of the acceptable values for that field.

Note: The schema is case-sensitive and will sometimes flag an otherwise correct entry if the capitalization does not match what is in the standard exactly. These are not actually errors and can either be ignored or capitalized to match the standard.

5. Apply a style-sheet to view the metadata in a more easily read format.

At any time in this process you can view the metadata in a more human-readable format by switching to the XSL Output tab and specifying a display style-sheet. To do this, click on the button with the three dots and navigate to a XSL style-sheet (this will be a file with an .xsl file extension). Then click the Transform button to change the display. It's not possible to make any edits while in this view, but this is a convenient way to review a record after editing it.
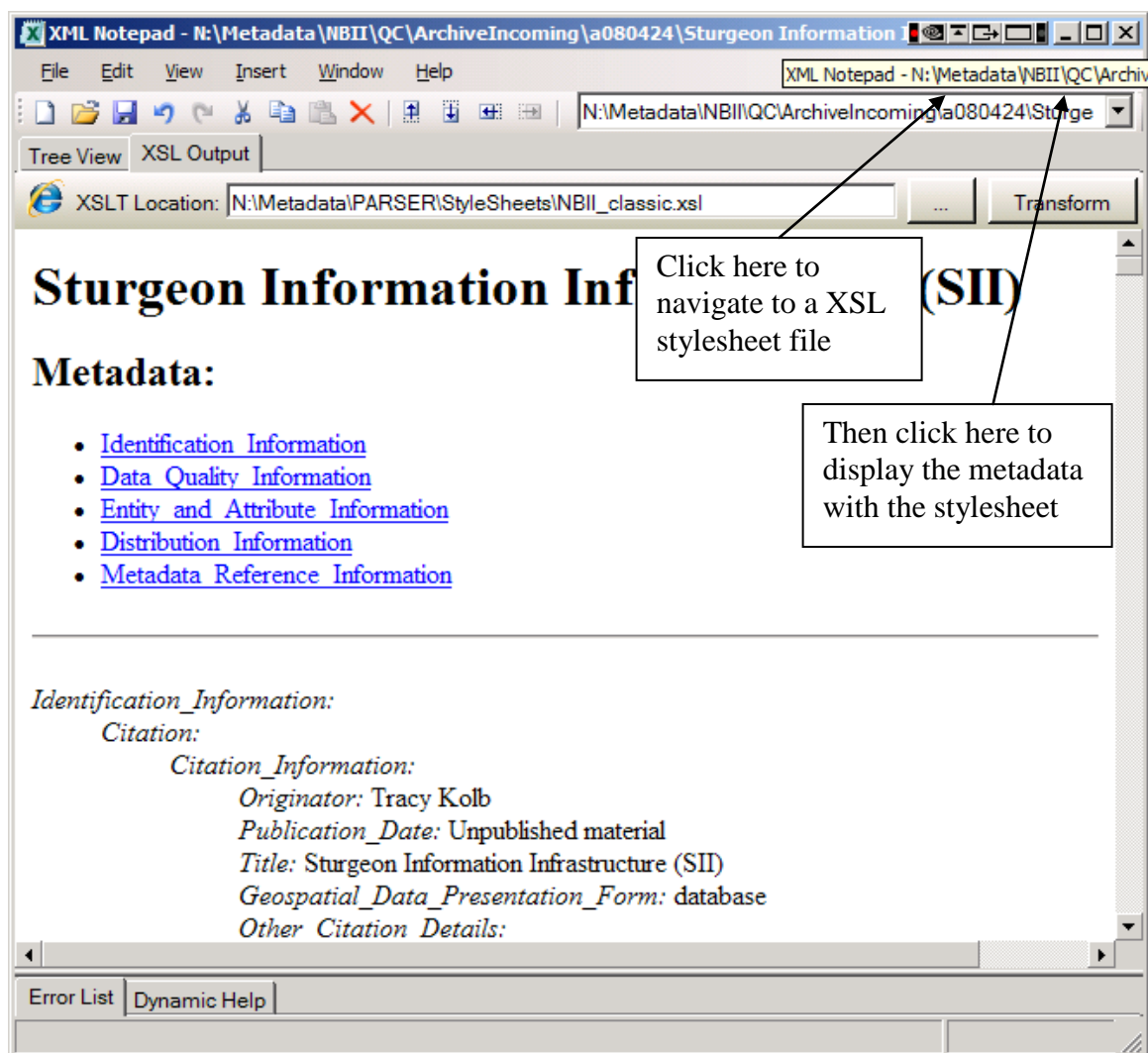
Appendix 1. – Applying/Reconfiguring Schemas in XML Notepad

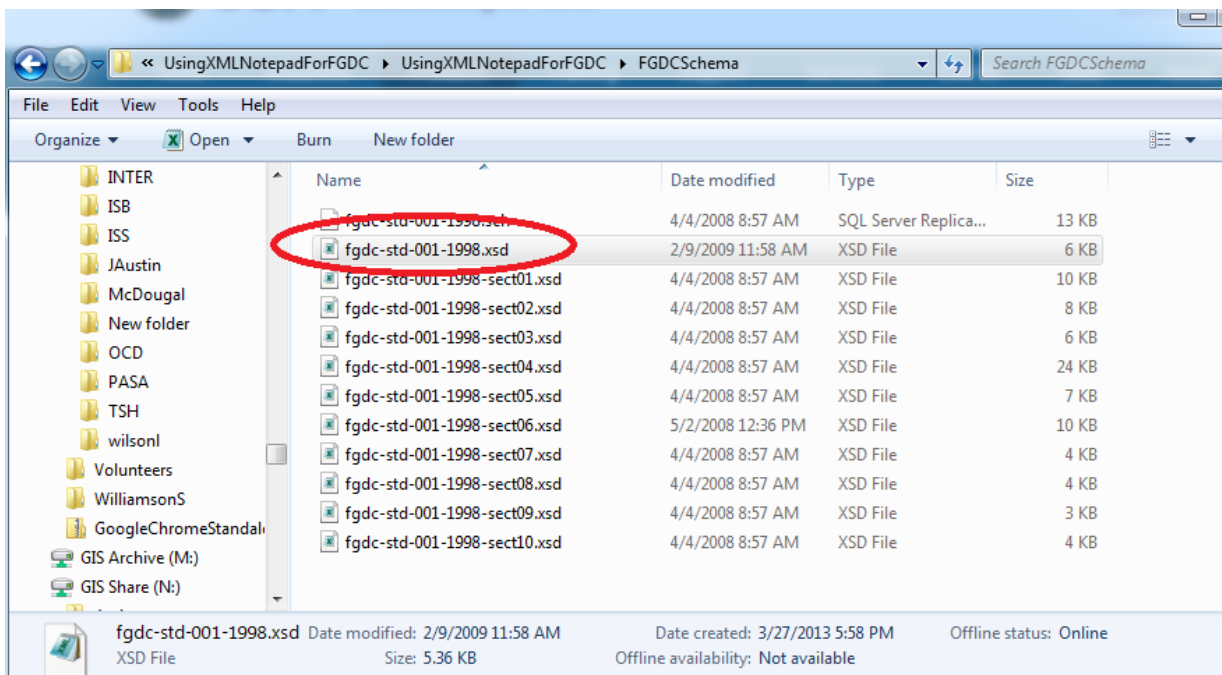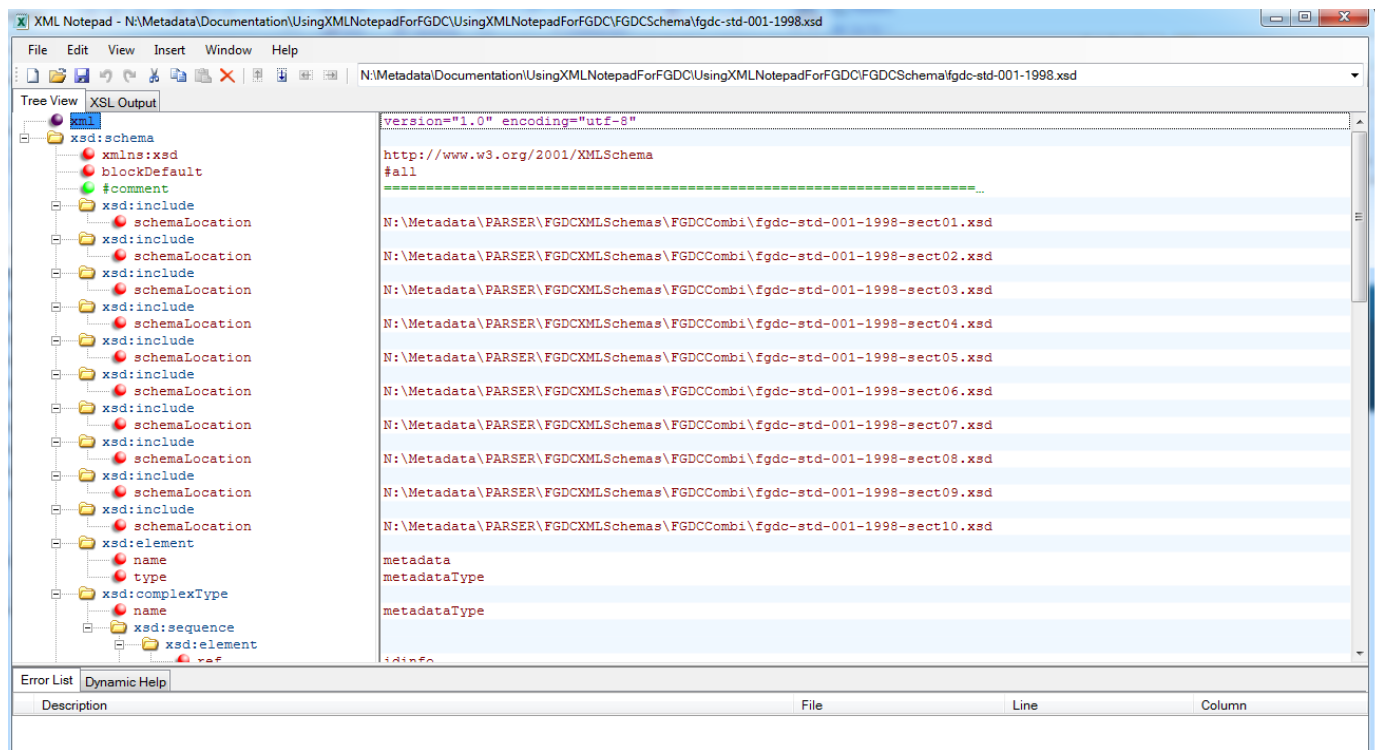Open the location where the XML schema materials are stored.
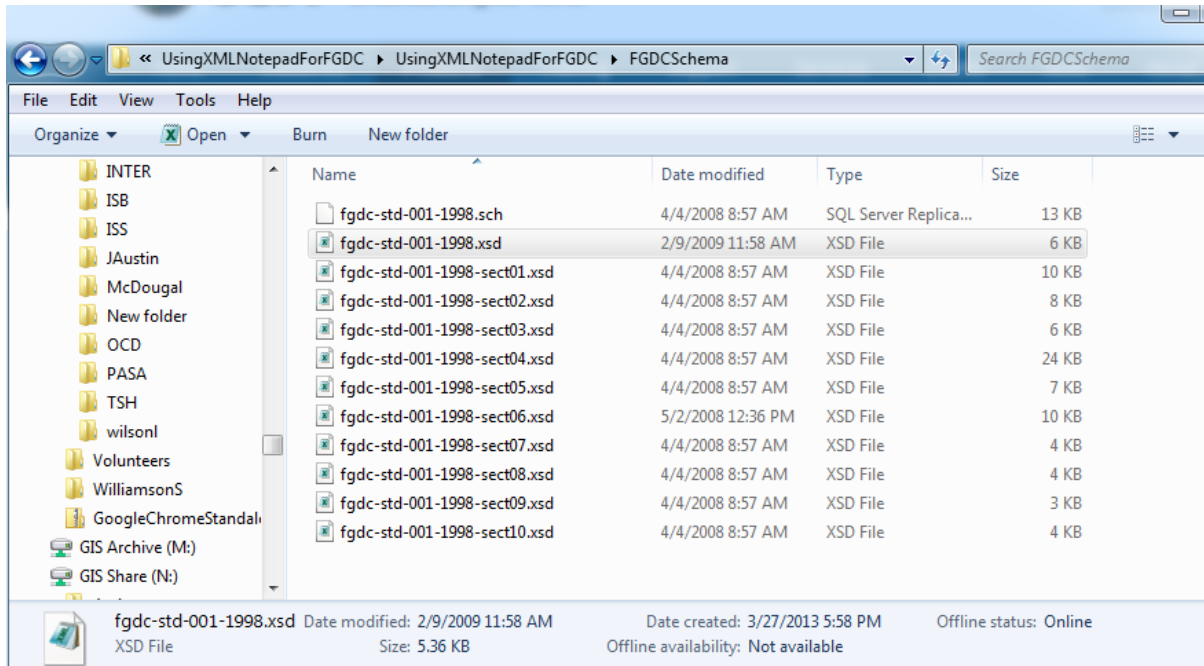
Open the first .xsd file. If it does not open automatically in XML Notepad, right click the file and 'Open With' XML Notepad.

In XML Notepad, with the 'fgdc-std-001-1998.xsd' file open, click View, Expand All.

Most often, the 'unable to resolve schema' error is related to the fact that the schema materials are not being referenced properly. Bear in mind that the 'top-level' .xsd file needs to properly reference all of the 10 required other .xsd files for FGDC-CSDGM validation. These files need to be listed, with correct file paths (locations) to work properly.
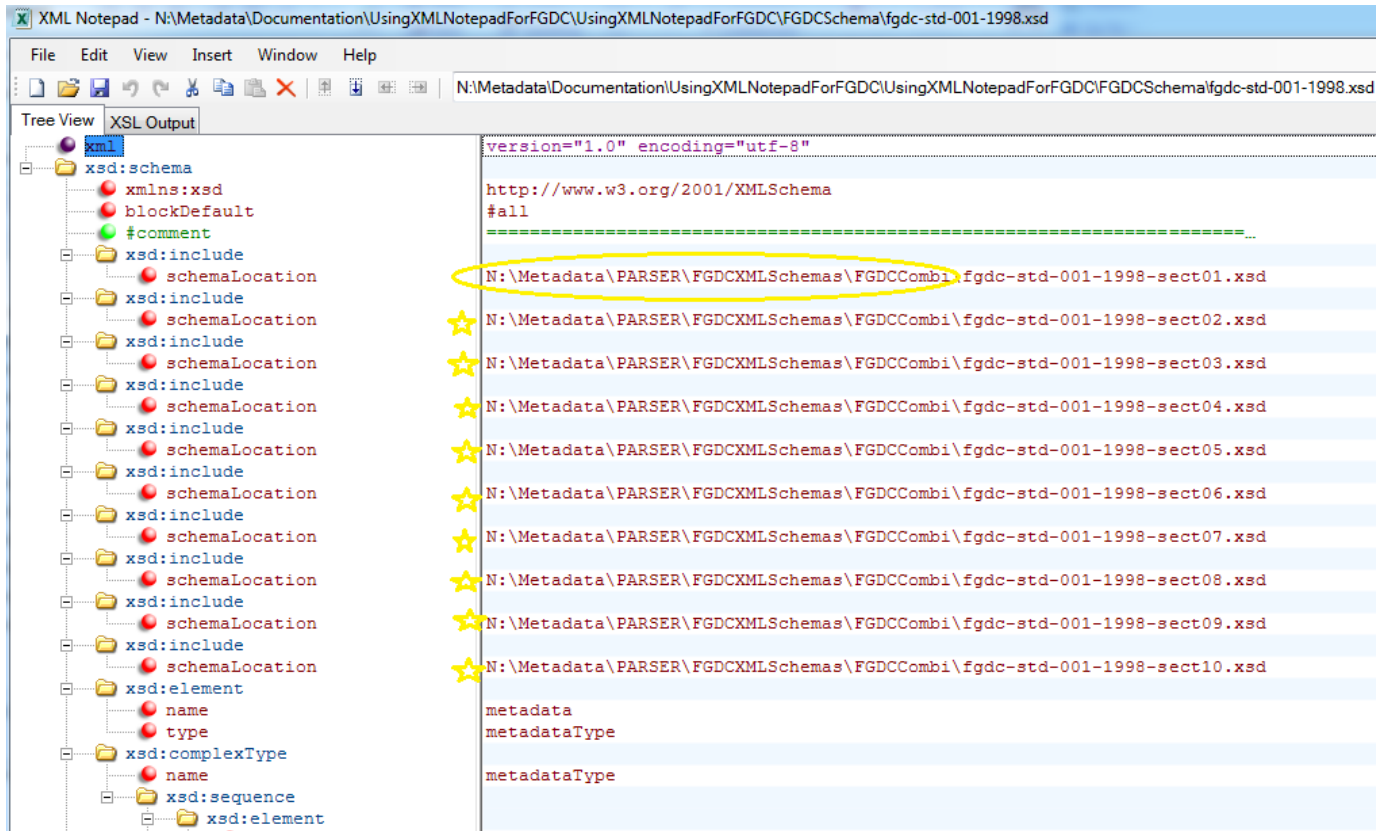
The above example shows what this looks like.

However if these paths are not current or the files are not present, XML Notepad will look for the other components of the schema in a location that does not exist on a user's system, and thus their 'location' can't be resolved.

If the 'unable to resolve schema location' is encountered, try editing each of the lines shown above to accurately reflect the location of these files. In the above example, this would be as simple as replacing the "N:\Metadata\PARSER\FGDCXMLSchemas\FGDCCombi" part of the path with the updated path to the folder where the schema materials reside (the folder containing all the .xsd files).

Save your changes.

Now, in XML Notepad, try clearing the schema again, restarting the program, View>Schemas and pointing to the first .xsd file again (shown below):

Name	Date modified	Type	Size
fgdc-std-001-1998.sch	4/4/2008 8:57 AM	SQL Server Replica...	13 KB
fgdc-std-001-1998.xsd	2/9/2009 11:58 AM	XSD File	6 KB
fgdc-std-001-1998-sect01.xsd	4/4/2008 8:57 AM	XSD File	10 KB
fgdc-std-001-1998-sect02.xsd	4/4/2008 8:57 AM	XSD File	8 KB
fgdc-std-001-1998-sect03.xsd	4/4/2008 8:57 AM	XSD File	6 KB
fgdc-std-001-1998-sect04.xsd	4/4/2008 8:57 AM	XSD File	24 KB
fgdc-std-001-1998-sect05.xsd	4/4/2008 8:57 AM	XSD File	7 KB
fgdc-std-001-1998-sect06.xsd	5/2/2008 12:36 PM	XSD File	10 KB
fgdc-std-001-1998-sect07.xsd	4/4/2008 8:57 AM	XSD File	4 KB
fgdc-std-001-1998-sect08.xsd	4/4/2008 8:57 AM	XSD File	4 KB
fgdc-std-001-1998-sect09.xsd	4/4/2008 8:57 AM	XSD File	3 KB
fgdc-std-001-1998-sect10.xsd	4/4/2008 8:57 AM	XSD File	4 KB

At this point, the schema should be accurately reflecting the proper path to the various .xsds that constitute the schema.

Once the schema is properly configured, dynamic error checking against the XML schema should be enabled. This will provide a user with up-to-date information about the compliance of a given file's organization and content with the FGDC-CSDGM metadata standard.